

Graphlet characteristics in directed networks

Igor Trpevski¹, Tamara Dimitrova¹, Tommy Boshkovski¹, and Ljupco Kocarev^{1,2,3,*}

¹Macedonian Academy of Sciences and Arts, Skopje, Republic of Macedonia

²Faculty of Computer Science and Engineering, UKIM, Skopje, Republic of Macedonia

³BioCircuits Institute, UC San Diego, La Jolla, CA 92093-0402, USA

March 23, 2016

A number of network structural characteristics have recently been the subject of particularly intense research, including degree distributions [1], community structure [2, 3], and various measures of vertex centrality [4, 5], to mention only a few. Vertices may have attributes associated with them; for example, properties of proteins in protein-protein interaction networks [6], users' social network profiles [7], or authors' publication histories in co-authorship networks [8]. In a network, two vertices might be considered similar if they have similar attributes (features, properties), or they can be considered similar based solely on the network structure. Similarity of this type is called structural similarity, to distinguish it from properties similarity, social similarity, textual similarity, functional similarity or other similarity types found in networks. Here we focus on the similarity problem by computing (1) for each vertex a vector of structural features, called *signature vector*, based on the number of graphlets associated with the vertex, and (2) for the network its *graphlet correlation matrix*, measuring graphlets dependencies and hence revealing unknown organizational principles of the network. Real-world networks are directed and often have few types of vertices with well defined structural (topological) characteristics. We found that real-world networks generally have very different structural characteristics resulting in different graphlet correlation matrices. In particular, the graphlet correlation matrix of the brain effective network is computed for 40 healthy subjects and common (present in more than 70% subjects) dependencies are revealed. Thus, negative correlations are found for 2-node graphlets and 3-node graphlets that are wedges and positive correlations are found only for 3-node graphlets that are triangles. Graphlets characteristics in directed networks could further significantly increase our understanding of real-world networks.

Quantifying similarity between vertices in a network is an old problem. Researches have proposed comparable vertex similarity measures predicating on the idea that vertices are similar if their neighbors are similar. In social network analysis, Katz centrality [9] of a vertex equals to the sum of that vertex's similarities to every other vertex and is computed by taking into account the total number of walks between a pair of vertices. In computer science, Jeh and Widom [10] have proposed a method that they call "SimRank," implementing a recursive definition of object similarity: two objects are similar if they relate to similar objects. SimRank similarity of two nodes i and j has an elegant random walk interpretation as the probability that two independent simultaneous random walkers, beginning at i and j , will eventually meet at some node. In physics, another similar concept which led to a self-consistent matrix formulation of similarity that can be evaluated iteratively using only the knowledge of the graph adjacency matrix is proposed by Newman and co-workers [11].

Here we measure structural similarity between vertices on directed network by considering *directed graphlets* (induced sub-graphs) associated with vertices, and, subsequently probe the structure of graphs by computing correlations of the frequencies of different graphlets present in the graph. Let $G = (V, E)$ be a simple directed (unweighted) graph and let $\mathbf{A} = [a_{ij}]$ be its $n \times n$ adjacency matrix, where n is the number of vertices, such that $a_{ii} = 0$ (self-arcs are excluded). For an element of the edge set E we write (i, j) to denote $i \rightarrow j$. A graphlet $G_k = (V_k, E_k)$ is an induced sub-graph that consists of a subset of k vertices of the graph $G = (V, E)$ (i.e., $V_k \subset V$) together with all the edges whose endpoints are both in this subset (i.e., $E_k = \{e \in E : e = (u, v) \text{ and } u, v \in V_k\}$). Therefore, graphlets are small induced sub-graphs of a large network that appear at any frequency and hence are independent of a null model. Here, we consider only weakly connected graphlets.

Graphlets have found numerous applications as building blocks of network analysis in various disciplines ranging from social science [12, 13] to biology [14, 15, 16, 2]. In social science, graphlet analysis (known as sub-graph census) is widely adopted in sociometric studies [12]. Much of the work in this vein focused on analyzing triadic tendencies as important structural features of social networks (e.g., transitivity or triadic closure) as well as analyzing triadic configurations as the basis for various social network theories (e.g., social balance, strength of weak ties, stability of ties, or trust [18]). Directed graphs are treated as having two different types of edges: directed and reciprocal. The research on reciprocal edges originates with the triad census work of Holland and Leinhardt [12]. A reciprocal edge is technically a pair of directed edges, $\{(i, j), (j, i)\}$, that we treat as a single reciprocal edge. We define the set of in-neighbors $S_i^- = \{j : (j, i) \in E \text{ and } (i, j) \notin E\}$, the set of out-neighbors $S_i^+ = \{j : (i, j) \in E \text{ and } (j, i) \notin E\}$, and the set of reciprocal-neighbors $S_i^o = \{j : (i, j) \in E \text{ and } (j, i) \in E\}$.

For a node i of network G , the automorphism orbit of i is the set of nodes of G that can be mapped to i by an automorphism, an isomorphism of a network with itself; i.e., a bijection of nodes that preserves node adjacency. Automorphism orbits have been used to generalize the graph degree distribution [16, 2]. Given a family of graphlets of size k nodes, $2 \leq k \leq 4$, one can count the number of times a node *is touched* by an orbit. In directed graphs there are 1695 orbits for up to 4-node graphlets, see SI. The larger number of such orbits makes a complete enumeration unattractive, a problem which continues to worsen for larger graphlets. Therefore, here we restrict ourselves to working with 3-node graphlets, reducing the number of orbits to 48. In addition, we further reduce it to 16 orbits by considering only isomorphism triangle and wedge classes *starting* at vertex i (see SI).

Triangle is a set of three nodes, each of which is connected to the other two. Let $T = (V_t, E_t)$ be a triangle with nodes i, h, j and edges e_1, e_2, e_3 between nodes i and h , h and j , and i and j respectively. We equip the edges with an arbitrary orientation, as this is necessary for the further analysis. To be specific, we assume that directed or reciprocal character of the edges between i and h and i and j are considered with respect of the node i and of the edge between h and j with respect of the node j . Thus, for example, we write $(+, -, \circ)$ triangle to denote the triangle $T = (V_t, E_t)$ with $V_t = \{i, h, j\}$ and $E_t = \{i \rightarrow h, h \rightarrow j, i \leftrightarrow j\}$. More generally we write (α, β, γ) triangle to denote one of the triangle types, where $\alpha, \beta, \gamma = \{\circ, +, -\}$. In a similar fashion we denote (α, β) wedges for the wedges in directed graphs. Since directed graphs have two types of edges: directed and reciprocal, there exist 9 wedge types, which are arranged in 6 isomorphism wedge classes (Figure 1a), and 27 triangle types, which can be grouped in 7 isomorphism triangle classes (Figure 1b). Finally, the 2-node graphlets orbits originating at a node i are equivalent to the out-links, in-links and reciprocal links of node i .

The number of (α, β, γ) triangles *starting* at i is given by:

$$T_i(\alpha, \beta, \gamma) = \sum_{j \in S_i^\gamma} |S_i^\alpha \cap S_j^\beta| = \sum_{h, j; h \neq j \neq i} a_{\alpha} a_{\beta} a_{\gamma}$$

where the entries $a_{\alpha}, a_{\beta}, a_{\gamma}$ of the graph adjacency matrix are provided in Figure 1c. Indeed, $|S_i^\alpha \cap S_j^\beta|$ is the number of the common α -neighbors of i and β -neighbors of j . Summarizing $|S_i^\alpha \cap S_j^\beta|$ for all j that are γ -neighbors of i , one computes the number of (α, β, γ) triangles. The number of (α, β) wedges starting at vertex i is $L_i^{(\alpha, \beta)} = \sum_j |S_i^\alpha \cap S_j^\beta|$; the number of (α, β) wedges that are graphlets is given by $W_i^{(\alpha, \beta)} = L_i^{(\alpha, \beta)} - \sum_{\gamma} T_i^{(\alpha, \beta, \gamma)}$. Therefore 39 quantities $d_i^\alpha = |S_i^\alpha|$, $W_i^{\alpha, \beta}$, and $T_i^{\alpha, \beta, \gamma}$ are associated with a vertex i . This number is further reduced to 16 by considering only isomorphic wedge and triangle classes (SI). We define *signature/feature* vector of a vertex i as $F_i = [d_i, W_i, T_i]^T$ where:

$$\begin{aligned} d_i &= [d_i^+, d_i^-, d_i^o]^T \\ W_i &= [W_i^{(path)}, W_i^{(in)}, W_i^{(out)}, W_i^{(in+)}, W_i^{(out+)}, W_i^{(rec)}]^T \\ T_i &= [T_i^{(acyclic)}, T_i^{(cycles)}, T_i^{(out+)}, T_i^{(cycles+)}, T_i^{(in+)}, T_i^{(cycles++)}, T_i^{(rec)}]^T \end{aligned}$$

d_i^+, d_i^-, d_i^o are the numbers of 2-node graphlets, or equivalently, the out-degree, in-degree and reciprocal-edge degree of the vertex, while W_i and T_i , called *wedge-degree* and *triangle-degree*, respectively, are the number of 3-node graphlets *starting* at i . Signature vectors describe only local connectivity of directed edges. Indeed, by flipping

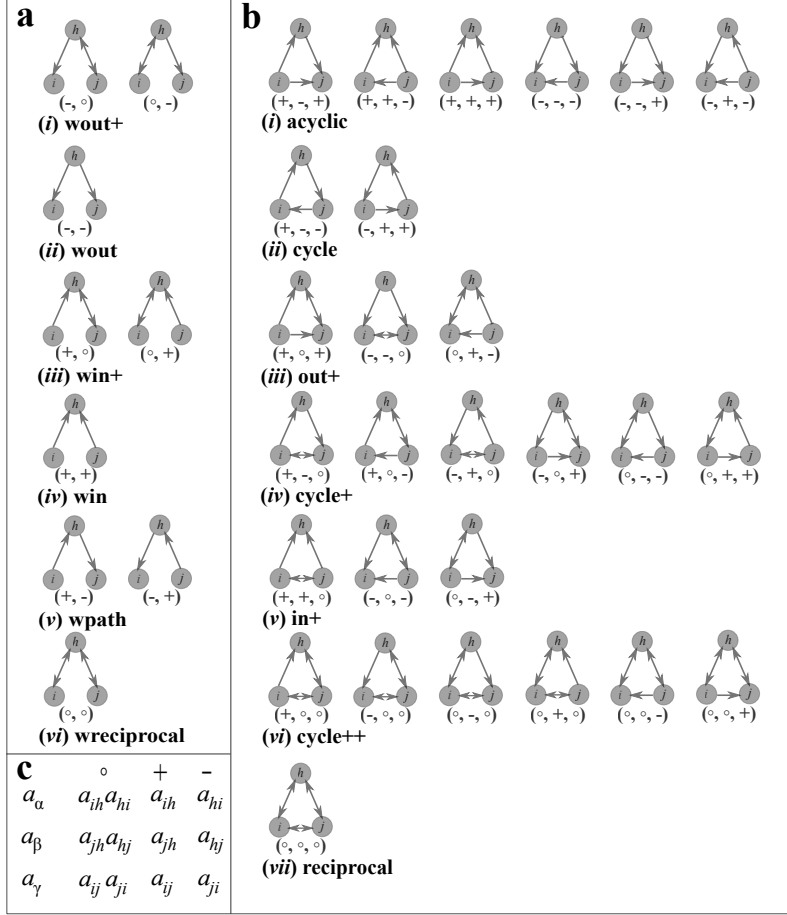


Figure 1: Isomorphic classes of wedges (a) and triangles (b) in directed graphs. In each wedge/triangle one vertex is labeled i (wedge/triangle starts at i). Assuming that directed (and reciprocal) edges are considered with respect to particular vertex in the wedge or the triangle (see the main text), each wedge and triangle can be labeled as (α, β) and (α, β, γ) , respectively, where $\alpha, \beta, \gamma \in \{+, -, \circ\}$. Hence, there are 9 wedges and 27 triangles starting at i , which are clustered in 6 wedge isomorphic classes (a) and 7 triangle isomorphic classes (b). (c) Entries of adjacency matrix for out-, in-, and reciprocal-edges.

the direction of edges between all connected pairs of nodes with probability one-third (for out-, in-, and reciprocal-edges), which keeps the undirected connectivity and thus the edge density unchanged, the information about the direction of edges will be destroyed and normalized signature vectors will converge to $[\frac{1}{3}, \frac{1}{3}, \frac{1}{3}, \frac{1}{6}, \dots, \frac{1}{6}, \frac{1}{7}, \dots, \frac{1}{7}]^T$ (see SI).

For each vertex in a network, we construct its signature vector consisting of 16 coordinates corresponding to the number of 16 graphlets starting at the vertex. Then we construct an $N \times 16$ matrix whose rows are the signature vectors for each vertex. For a given network G , we compute Pearson correlation coefficients between all pairs of columns of the above described matrix and present them in a 16×16 symmetric matrix that is termed [2] *graphlet correlation matrix* of network G . In this way, the network topology and its local direction patters, regardless network size (the number of vertices) and network volume (the number of edges), are summarized into a 16×16 matrix.

Different networks generally have very different graphlet dependencies and hence reveal unknown organizational principles of real-world networks. We study effective brain networks of 40 healthy subjects (see supplementary information). Brain connectivity refers to a pattern of anatomical links (“anatomical connectivity”), of statistical dependencies (“functional connectivity”) or of causal interactions (“effective connectivity”) between distinct

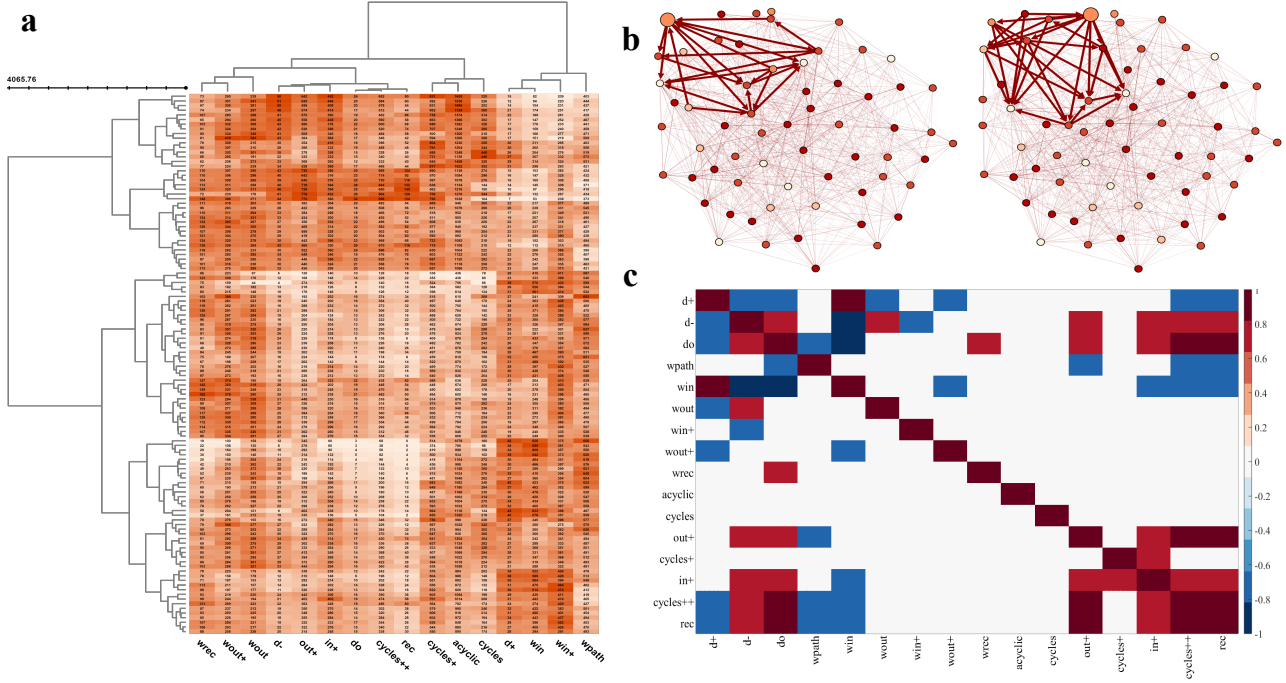


Figure 2: Graphlets in effective brain network. The dataset represents effective connectivity of the brain network, describing a network of directional effects of one neural region over another. (a) 116 regions (vertices) are considered in total (shown on the vertical axis); a directed edge represents causality of one region (vertex) over another. Each region is represented by 16-dimensional feature vector (shown on horizontal axis). Ward agglomerative hierarchical clustering procedure results in a heat-map representing clustered regions. Three clusters with similar local structure are easily identified. (b) Two regions from the same cluster are shown, both having similar local structure; only for better visualization, sub-graph with 64 nodes is shown on which the local structures around two (similar) vertices are shown. (c) Graphlet correlation matrix of a particular brain network is shown, only those entries for which correlations (or anti-correlations) are significant (> 0.7 and < -0.7) are colored.

units – region of interests (ROI) – within a nervous system [19]. Here, effective brain connectivity is studied using Granger (G-) causality. Resting-State fMRI Data Analysis Toolkit (REST) is used and multivariate-coefficients ROI-wise G - causality analysis (GCA) is adopted to generate effective networks. REST-GCA supports signed-path coefficients multivariate GCA and integrates a batch mode coefficients computation for ROI-wise GCA [20]. For each network the number of vertices (regions) is equal to 116, while the total number of edges in the networks (after pruning) is 3466 ± 43 . We construct a matrix such that the number of rows in the matrix is equal to the number of vertices in the network and the number of columns equals the dimension of the signature vector (the matrix has 16 columns), see Figure 2a. The signature vector represents the local structure around the vertex and can be employed to understand organizational rules within a network and to reveal structural differences between networks. We also compute correlation coefficients between all pairs of columns of this matrix and present them in a 16×16 symmetric matrix, Figure 2c.

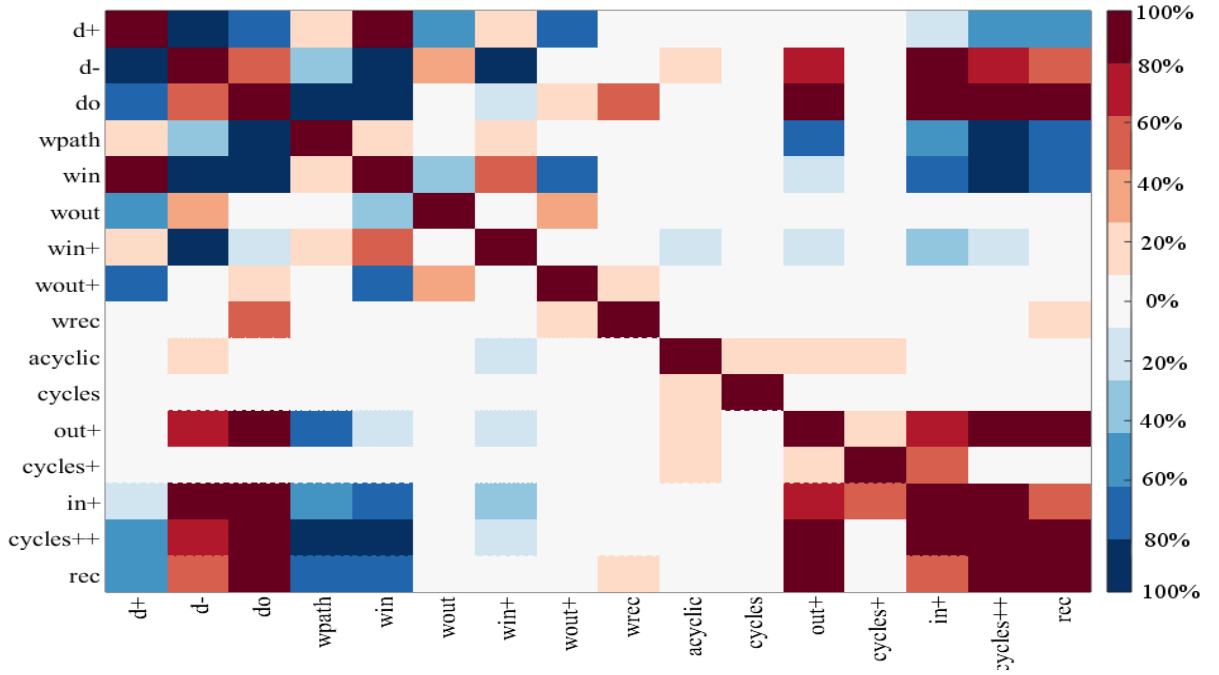


Figure 3: Graphlet correlation matrices are computed for all 40 healthy subjects. Percentages of the healthy subjects that have statistically significant are colored. The correlation is considered significant if the Pearson correlation coefficient is greater than 0.7 and for anti-correlation is significant if the coefficient is less than -0.7 . The heat-map indicates that there are many pairs of entries of the signature vector that are significantly correlated or anti-correlated for most of the subjects.

We then compute the percentage of subjects for which particular coordinates (similarity metrics) of the signature vectors are significantly correlated (> 0.7) or anti-correlated (< -0.7), Figure 3. We found that for brain effective networks higher values of the number of out-degree is associated (significantly anti-correlated) with lower values of the number of in-degree in 95% subjects and with the number of reciprocal-degree in 64% subjects. Therefore, regions which are involved in activations (G-causes) of other regions (higher values of G-causes manifested with higher values of d^+) at the same time are not activate (G-effects) by other regions (lower values of G-effects shown in both d^- and d°). On the other hand, the number of in-degree and the number of reciprocal-degree are correlated in 62% subjects. Moreover, we found that $W^{(path)}$ is anti-correlated with $T^{(out+)}$ (in 74% subjects), $T^{(cycle++)}$ (in 85% subjects) and with $T^{(rec)}$ (in 79% subjects). The number of in-wedges $W^{(in)}$ is anti-correlated

with $W^{(out+)}$ (in 74% subjects), $T^{(in+)}$ (in 79% subjects), $T^{(cycle++)}$ (in 85% subjects) and with $T^{(rec)}$ (in 74% subjects). While the relationships between degrees, wedge-degrees, and wedge-degrees and triangle-degrees are mostly negative (anti-correlated), we found that relationships between different number of triangles (involving reciprocal edges) are positive (correlated): (1) $T^{(out+)}$ with $T^{(in+)}$ (69%), $T^{(cycle++)}$ (100%), $T^{(rec)}$ (97%); (2) $T^{(cycle++)}$ with $T^{(in+)}$ (59%); (3) $T^{(in+)}$ with $T^{(cycle++)}$ (100%) and $T^{(rec)}$ (59%); (4) $T^{(cycle++)}$ with $T^{(rec)}$ (100%). However, no significant (positive or negative) relationships are found in which the number of acyclic triangles and the number of cycle triangles are involved.

In conclusion vertex signature vector and graphlet correlation matrix are powerful tools for network analysis. We found that the graphlet correlation matrix of the brain effective network has positive correlation between various triangle-degrees only if at least one edge is reciprocal. Negative correlation was pronounced when comparing out-degrees with in-degrees and reciprocal-degrees, as well as, when comparing degrees with wedge-degrees. How this finding will further enhance our understanding of brain connectivity will be a subject of our future research activities.

After submitting this paper we learned that Przulj's group was working on similar problem using different approach and data [21].

References

- [1] A.-L. Barabasi and R. Albert, "Emergence of Scaling in Random Networks," Science, Vol 286, Issue 5439, 509–512, 1999.
- [2] M. Girvan and M. E. J. Newman, "Community structure in social and biological networks," Proc. Natl. Acad. Sci. USA, vol. 99, no. 12, 7821–7826, 2002.
- [3] F. Radicchi, C. Castellano, F. Cecconi, V. Loreto and D. Parisi, "Defining and identifying communities in networks," Proc. Natl. Acad. Sci. USA 101, no. 9, 2658–2663, 2004.
- [4] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," Nature 393, 440–442, 1998.
- [5] F. Chung and L. Lu, "The average distances in random graphs with given expected degrees," Proc. Natl. Acad. Sci. USA, vol. 99, no. 25, 15879–15882, 2002.
- [6] Min-Sik Kim et al, "A draft map of the human proteome," Nature 509, 575–581, 2014.
- [7] B. Wellman, "Computer networks as social networks," Science Vol. 293, Issue 5537, 2031–2034, 2001.
- [8] M. E. J. Newman, "Coauthorship networks and patterns of scientific collaboration," Proc. Natl. Acad. Sci. USA, vol. 101, 5200–5205, 2004.
- [9] L. Katz, "A New Status Index Derived from Sociometric Index," Psychometrika, vol 18, 39–43, 1953.
- [10] G. Jeh and J. Widom, "Simrank: a measure of structural-context similarity," In Proc. 8th ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining, 538–543, New York, NY, 2002.
- [11] E. A. Leicht, P. Holme, and M. E. J. Newman, "Vertex similarity in networks," Phys. Rev. E 73, 026120, 2006.
- [12] P. W. Holland and S. Leinhardt, "Local structure in social networks," Sociological Methodology 7, 1–45, 1976.
- [13] K. Faust, "A puzzle concerning triads in social networks: Graph constraints and the triad census," Social Networks, vol. 32, 221–233, 2010.
- [14] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon, "Network motifs: simple building blocks of complex networks," Science 298, 824, 2002.
- [15] O. Sporns and R. Kötter, "Motifs in Brain Networks," PLoS Biol 2(11): e369, 2004.

- [16] N. Pržulj, “Biological network comparison using graphlet degree distribution,” *Bioinformatics* 23, 177–183, 2007.
- [17] O. N. Yaveroğlu, N. Malod-Dognin, D. Davis, Z. Levnajic, V. Janjic, R. Karapandza, A. Stojmirovic & N. Pržulj, “Revealing the Hidden Language of Complex Networks,” *Scientific Reports* 4, Article number: 4547, 2014.
- [18] M. Granovetter, “The strength of weak ties: A network theory revisited,” *Sociological theory*, vol. 1(1), 201–233, 1983.
- [19] E Bullmore and O Sporns, “Complex brain networks: graph theoretical analysis of structural and functional systems,” *Nature Reviews Neuroscience*, vol. 10, no. 3, 186–198, 2009.
- [20] Xiao-Wei Song, Zhang-Ye Dong, Xiang-Yu Long, Su-Fang Li, Xi-Nian Zuo, Chao-Zhe Zhu, Yong He, Chao-Gan Yan, Yu-Feng Zang, “REST: A Toolkit for Resting-State Functional Magnetic Resonance Imaging Data Processing”, *PLoS ONE* 6(9): e25031, 2011.
- [21] <http://arxiv.org/pdf/1603.05470v1.pdf>

Supplementary Information to the paper “Graphlet characteristics in directed networks”

Here we give detailed explanations and provide additional examples to the main paper [1].

1 Graphlet-based similarity

In undirected graphs, there are eight 2- to 4-node graphlets resulting in 15 orbits [2] (of which 11 are non-redundant [2]), that touch a vertex i . Since the number of up to 4-mode graphlets in directed networks is large, here we consider only a subset of these orbits starting (originating) from a vertex. More specifically, let h_1, \dots, h_k , be nodes belonging to graphlets with $2 \leq k \leq 4$ nodes and define $\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6 \in \{+, -, \circ\}$. We enumerate all possible ways in which an orbit can touch the node h_1 in these directed graphlets:

$$\begin{aligned}
 O_0(h_1)^{(\alpha_1)} &= S_{h_1}^{\alpha_1} \\
 O_1(h_1)^{(\alpha_1, \alpha_2)} &= \{h_2, h_3 : h_2 \in S_{h_1}^{\alpha_1}, h_3 \in S_{h_2}^{\alpha_2}\} \\
 O_2(h_1)^{(\alpha_1, \alpha_2)} &= \{h_2, h_3 : h_2 \in S_{h_1}^{\alpha_1}, h_3 \in S_{h_1}^{\alpha_2}\} \\
 O_3(h_1)^{(\alpha_1, \alpha_2, \alpha_3)} &= \{h_2, h_3 : h_2 \in S_{h_1}^{\alpha_1}, h_3 \in S_{h_1}^{\alpha_2} \cap S_{h_2}^{\alpha_3}\} \\
 O_4(h_1)^{(\alpha_1, \alpha_2, \alpha_3)} &= \{h_2, h_3, h_4 : h_{i+1} \in S_{h_i}^{\alpha_i}, i = 1, 2, 3\} \\
 O_5(h_1)^{(\alpha_1, \alpha_2, \alpha_3)} &= \{h_2, h_3, h_4 : h_2 \in S_{h_1}^{\alpha_1}, h_3 \in S_{h_1}^{\alpha_2}, h_4 \in S_{h_3}^{\alpha_3}\} \\
 O_6(h_1)^{(\alpha_1, \alpha_2, \alpha_3)} &= \{h_2, h_3, h_4 : h_2 \in S_{h_1}^{\alpha_1}, h_3 \in S_{h_2}^{\alpha_2}, h_4 \in S_{h_2}^{\alpha_3}\} \\
 O_7(h_1)^{(\alpha_1, \alpha_2, \alpha_3)} &= \{h_2, h_3, h_4 : h_i \in S_{h_1}^{\alpha_{i-1}}, i = 2, 3, 4\} \\
 O_8(h_1)^{(\alpha_1, \alpha_2, \alpha_3, \alpha_4)} &= \{h_2, h_3, h_4 : h_{i+1} \in S_{h_i}^{\alpha_i}, i = 1, 2, 3; h_4 \in S_{h_1}^{\alpha_4}\} \\
 O_9(h_1)^{(\alpha_1, \alpha_2, \alpha_3, \alpha_4)} &= \{h_2, h_3, h_4 : h_{i+1} \in S_{h_i}^{\alpha_i}, i = 1, 2, 3; h_4 \in S_{h_2}^{\alpha_4}\} \\
 O_{10}(h_1)^{(\alpha_1, \alpha_2, \alpha_3, \alpha_4)} &= \{h_2, h_3, h_4 : h_{i+1} \in S_{h_i}^{\alpha_i}, i = 1, 2, 3; h_3 \in S_{h_1}^{\alpha_4}\} \\
 O_{11}(h_1)^{(\alpha_1, \alpha_2, \alpha_3, \alpha_4)} &= \{h_2, h_3, h_4 : h_i \in S_{h_1}^{\alpha_{i-1}}, i = 2, 3, 4; h_4 \in S_{h_3}^{\alpha_4}\} \\
 O_{12}(h_1)^{(\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5)} &= \{h_2, h_3, h_4 : h_{i+1} \in S_{h_i}^{\alpha_i}, i = 1, 2; h_4 \in S_{h_i}^{\alpha_{i+2}}, i = 1, 2, 3\} \\
 O_{13}(h_1)^{(\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5)} &= \{h_2, h_3, h_4 : h_2 \in S_{h_1}^{\alpha_1}, h_3 \in S_{h_1}^{\alpha_2}; h_4 \in S_{h_i}^{\alpha_{i+2}}, i = 1, 2, 3\} \\
 O_{14}(h_1)^{(\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6)} &= \{h_2, h_3, h_4 : h_2 \in S_{h_1}^{\alpha_1}, h_3 \in S_{h_1}^{\alpha_2} \cap S_{h_2}^{\alpha_3}, h_4 \in S_{h_1}^{\alpha_4} \cap S_{h_2}^{\alpha_5} \cap S_{h_3}^{\alpha_6}\}
 \end{aligned}$$

resulting in a total of $1695 = 3 + 2 \times 3^2 + 5 \times 3^3 + 4 \times 3^4 + 2 \times 3^5 + 3^6$ combinations.

Considering only 3-node graphlets and focusing only on orbits which *originate* at node i , we define the following 39 quantities for a node i in the graph:

$$\begin{aligned}
 d_i^\alpha &= |S_i^\alpha| \\
 T_i^{(\alpha, \beta, \gamma)} &= \sum_{j \in S_i^\gamma} |S_i^\alpha \cap S_j^\beta| \\
 W_i^{(\alpha, \beta)} &= \sum_{j \neq i} |S_i^\alpha \cap S_j^\beta| - \sum_\gamma T_i^{(\alpha, \beta, \gamma)}
 \end{aligned}$$

These 39 quantities can be aggregated into 16 quantities by grouping wedges in 6 wedge isomorphic classes and triangles in 7 triangle isomorphic classes, resulting in:

- 3 degrees: d_i^+ , d_i^- , and d_i° ;
- 6 wedge-degrees:

$$\begin{aligned}
W_i^{(path)} &= W_i^{(+,-)} + W_i^{(-,+)} \\
W_i^{(in)} &= W_i^{(+,+)} \\
W_i^{(out)} &= W_i^{(-,-)} \\
W_i^{(in+)} &= W_i^{(+,\circ)} + W_i^{(\circ,+)} \\
W_i^{(out+)} &= W_i^{(-,\circ)} + W_i^{(\circ,-)} \\
W_i^{(rec)} &= W_i^{(\circ,\circ)};
\end{aligned}$$

- 7 triangle-degrees:

$$\begin{aligned}
T_i^{(acyclic)} &= T_i^{(+,-,+)} + T_i^{(+,+, -)} + T_i^{(+,+, +)} + T_i^{(-,-,-)} + T_i^{(-,-,+)} + T_i^{(-,+, -)} \\
T_i^{(cycles)} &= T_i^{(+,-,-)} + T_i^{(-,+, +)} \\
T_i^{(out+)} &= T_i^{(+,\circ,+)} + T_i^{(-,-,\circ)} + T_i^{(\circ,+, -)} \\
T_i^{(cycles+)} &= T_i^{(+,-,\circ)} + T_i^{(+,\circ,-)} + T_i^{(-,+, \circ)} + T_i^{(-,\circ,+)} + T_i^{(\circ,-,-)} + T_i^{(\circ,+, +)} \\
T_i^{(in+)} &= T_i^{(+,+, \circ)} + T_i^{(-,\circ,-)} + T_i^{(\circ,-,+)} \\
T_i^{(cycle++)} &= T_i^{(+,\circ,\circ)} + T_i^{(-,\circ,\circ)} + T_i^{(\circ,-,\circ)} + T_i^{(\circ,+, \circ)} + T_i^{(\circ,\circ,-)} + T_i^{(\circ,\circ,+)} \\
T_i^{(rec)} &= T_i^{(\circ,\circ,\circ)}.
\end{aligned}$$

The quantities degrees, wedge-degrees, and triangle-degrees could be normalized in different ways. For example, one can consider normalized signature vector defined as:

$$\bar{F}_i = \left[\frac{d_i^+}{A_i}, \frac{d_i^-}{A_i}, \frac{d_i^\circ}{A_i}, \frac{W_i^{(path)}}{B_i}, \dots, \frac{W_i^{(rec)}}{B_i}, \frac{T_i^{(acyclic)}}{C_i}, \dots, \frac{T_i^{(rec)}}{C_i} \right]^T$$

where

$$\begin{aligned}
A_i &= d_i^+ + d_i^- + d_i^\circ \\
B_i &= W_i^{(path)} + W_i^{(in)} + W_i^{(out)} + W_i^{(in+)} + W_i^{(out+)} + W_i^{(rec)}, \\
C_i &= T_i^{(acyclic)} + T_i^{(cycles)} + T_i^{(out+)} + T_i^{(cycles+)} + T_i^{(in+)} + T_i^{(cycles++)} + T_i^{(rec)}
\end{aligned}$$

By flipping the direction of edges between all connected pairs of nodes with probability one-third (for out-, in-, and reciprocal-edges), which keeps the undirected connectivity and the edge density unchanged, information about the direction of edges will be destroyed and normalized signature vectors should be close to $[\frac{1}{3}, \frac{1}{3}, \frac{1}{3}, \frac{1}{6}, \dots, \frac{1}{6}, \frac{1}{7}, \dots, \frac{1}{7}]^T$. This could be further exploited by studying normalized signature vectors for different real-world networks and their distances to random directed network with no direction-generated structure (such graph could be generated by considering arbitrary undirected connected graph and then transforming it to directed graph by assigning to each edge of the graph a direction (out, in, or reciprocal) with probability 1/3).

The number of triangles (but also wedges) could also be normalized as follows. $t_i(\alpha, \beta, \gamma)$ is the normalized number of (α, β, γ) triangles associated with the actor i :

$$t_i(\alpha, \beta, \gamma) = \frac{\sum_{j \in S_i^\gamma} |S_i^\alpha \cap S_j^\beta|}{\sum_{j \neq i} |S_i^\alpha \cap S_j^\beta|} = \frac{\sum_j x_\gamma \sum_h x_\alpha x_\beta}{\sum_j \sum_h x_\alpha x_\beta}$$

$\sum_{j \neq i} |S_i^\alpha \cap S_j^\beta| = \sum_{j \neq i} \sum_h x_\alpha x_\beta$ is the number of all (α, β) wedges starting at i . Thus, $t_i(\alpha, \beta, \gamma)$ indicates the ratio of (α, β) wedges that are (α, β, γ) triangles. This is a generalization of the concept of clustering coefficient of a graph to clustering coefficient of (α, β, γ) triangle associated to the node i :

$$t_i = \frac{\text{Number of } (\alpha, \beta, \gamma) \text{ triangles}}{\text{Number of } (\alpha, \beta) \text{ wedges}}$$

2 Effective brain networks

2.1 Data description

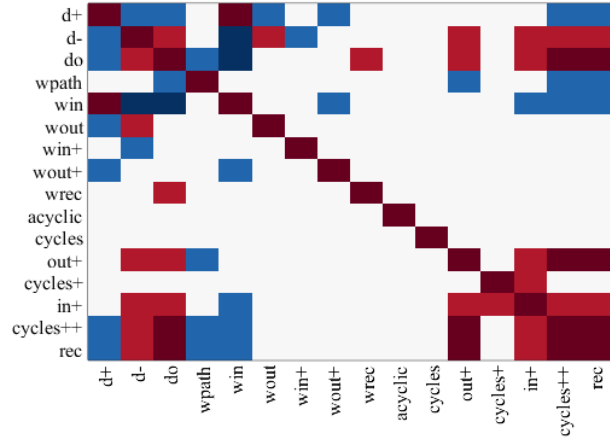
The effective brain networks were extracted from the rs - fMRI images of a group of healthy subjects and were taken from the Beijing Enhanced data set for multimodal brain images [3]. In this study, a total of 40 healthy subjects were used, aged from 18 to 26 years. The fMRI scans were taken during resting state of the subject, with a SIMENS Trio Tim scanner (whose strength of magnetics field of 3 Tesla, and EPI sequence). The following parameters of the EPI sequence were used: repetition (TR) / time echo (TE) = 2000 / 30 ms, flip angle = 90, field of view (FOV) = $200 \times 200 \text{ mm}^2$, voxel dimension = $3 \times 3 \times 3 \text{ mm}^3$, number of slices = 240. For preprocessing the fMRI data the DPARSF toolbox [4] was used. First, the images of each subject were corrected for slice timing and realigned (motion corrected). Then the images were normalized using the MNI EPI template with affine registration followed by nonlinear transformation. After, smoothing was applied using a Gaussian Kernel of 4mm Full Width at Half Maximum and the signal was detrended to remove any noise that may have remained from the previous steps. Finally, the data were filtered to preserve low frequency fluctuations (0.01 – 0.08 Hz). The effective connectivity for each subject were reconstructed using Granger causality coefficient between the time series of 116 regions of interest (ROIs). After obtaining effective connectivity for each subject, the weak connections from the effective connectivity network were pruned using threshold value, independently for each subject using. Using a pruning threshold must meet the following two criteria [5, 6, 7] : (i) 99 percent of the vertices (regions) are connected after pruning and (ii) the minimum degree of each vertex is at least $2 \times \ln(N)$, where N is the number of nodes in the network according to Network based statistic toolbox [8].

2.2 Graphlet correlation matrix

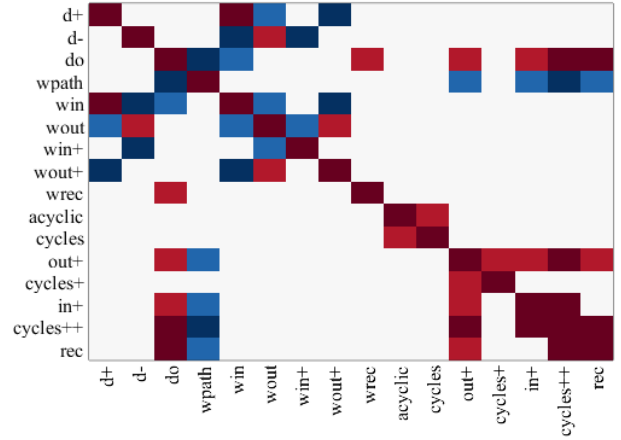
Here graphlet correlation matrices are provided for all 40 healthy subjects (Figure 1 to Figure 7).

References

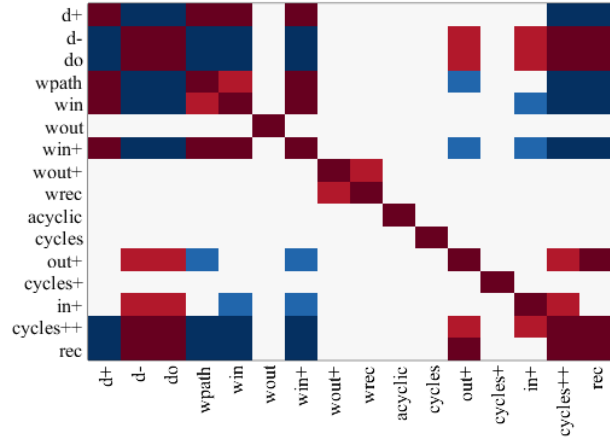
- [1] I.Trpevski, T.Dimitrova, T.Boshkovski and L.Kocarev, “Vertex signature vector for directed networks”
- [2] O. N. Yaveroğlu, N. Malod-Dognin, D. Davis, Z. Levnajic, V. Janjic, R. Karapandza, A. Stojmirovic & N. Pržulj, “Revealing the Hidden Language of Complex Networks”, Scientific Reports 4, Article number: 4547 (2014) doi:10.1038/srep04547
- [3] http://fcon_1000.projects.nitrc.org/indi/retro/BeijingEnhanced.html
- [4] Chao-Gan, Y., & Yu-Feng, Z. (2010). “DPARSF: A MATLAB Toolbox for "Pipeline" Data Analysis of Resting-State fMRI.”, Frontiers in Systems Neuroscience, 4, 13.
- [5] D. Papo, M. Zanin, and J. Buldú, “Reconstructing functional brain networks: have we got the basics right?”, Front. Hum. Neurosci., 2014.
- [6] Bassett, D. S., Meyer-Lindenberg, A., Achard, S., Duke, T., & Bullmore, E. (2006). “Adaptive reconfiguration of fractal small-world human brain functional networks.” Proceedings of the National Academy of Sciences, 103(51), 19518-19523.



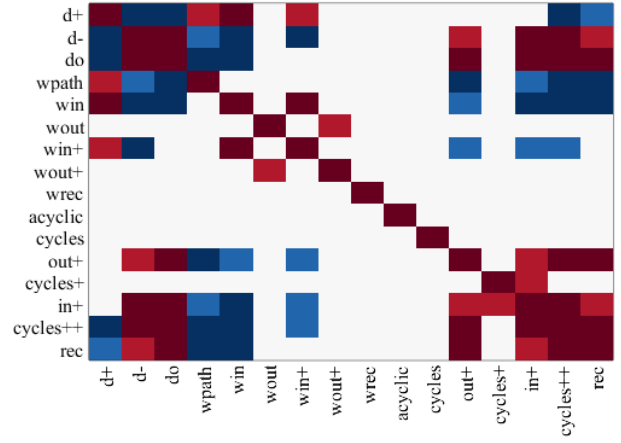
(a) Subject 1



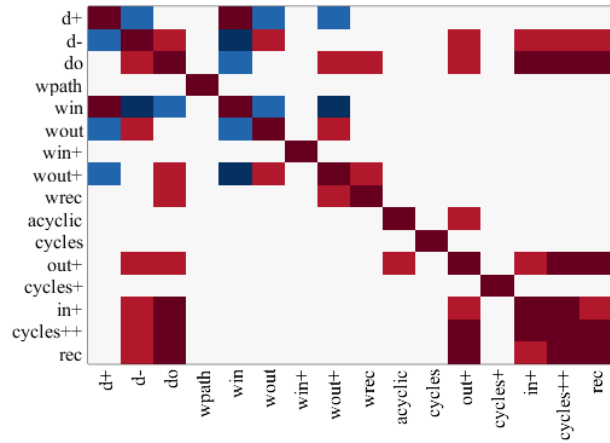
(d) Subject 4



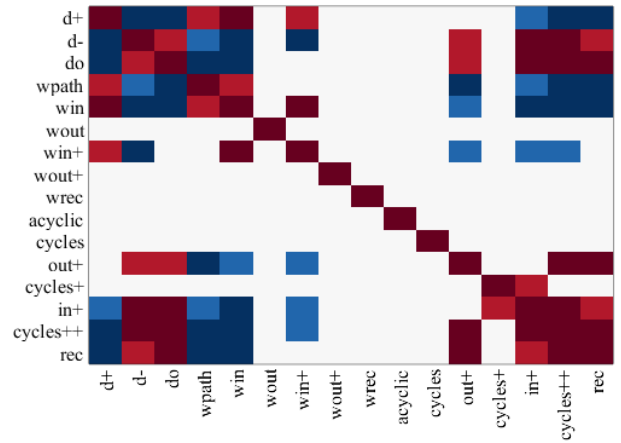
(b) Subject 2



(e) Subject 5

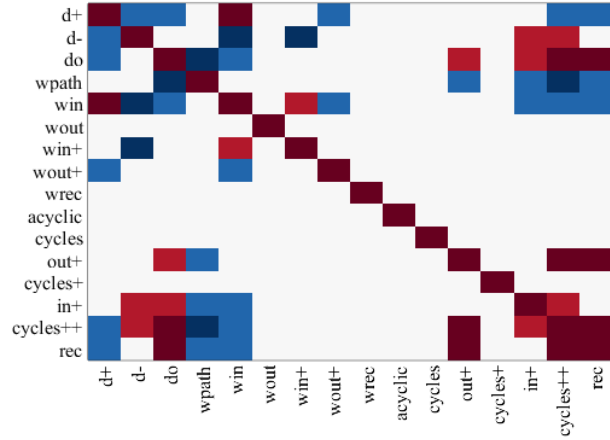


(c) Subject 3

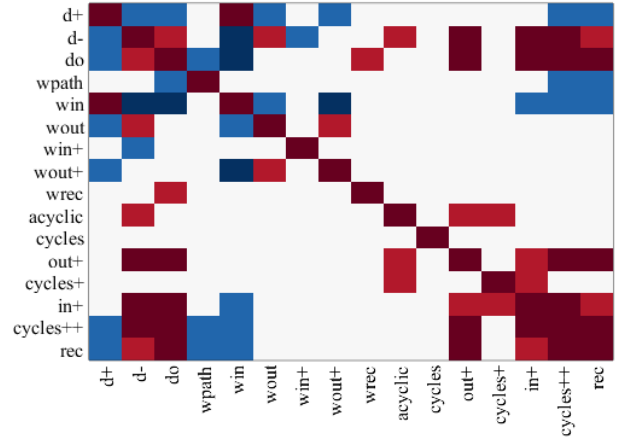


(f) Subject 6

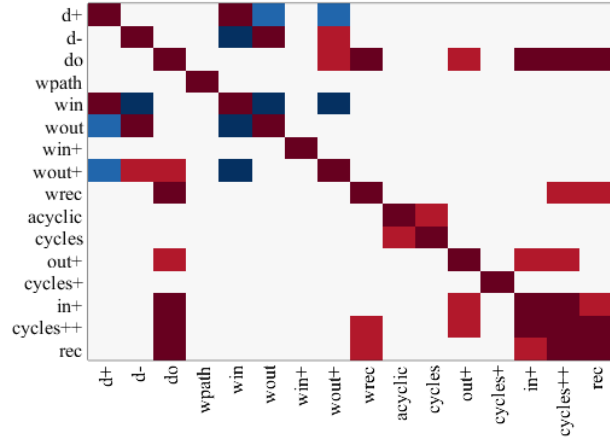
Figure 4: Graphlet correlation matrices for subjects 1 to 6



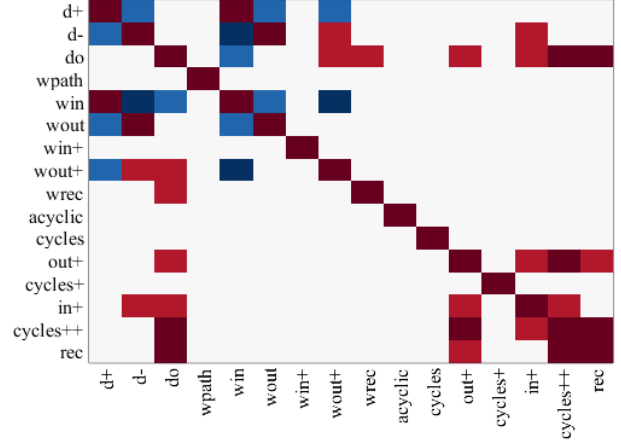
(a) Subject 7



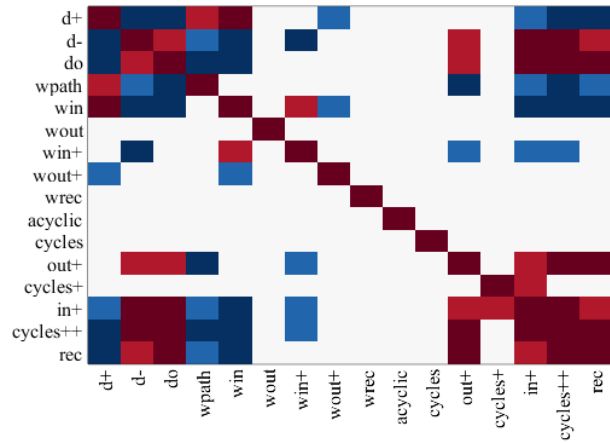
(d) Subject 10



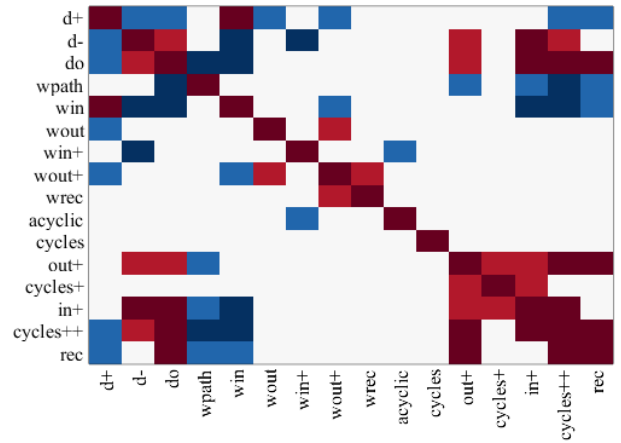
(b) Subject 8



(e) Patient 11

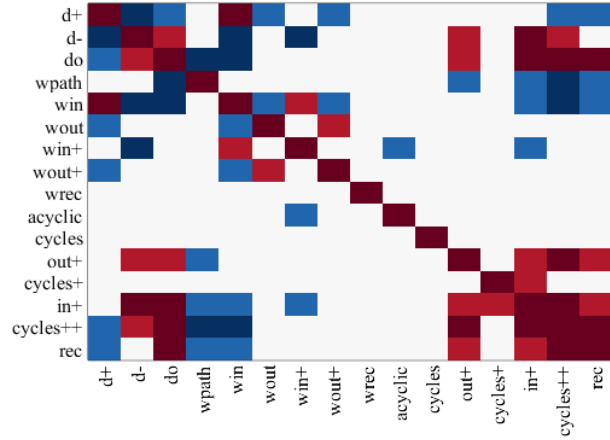


(c) Subject 9

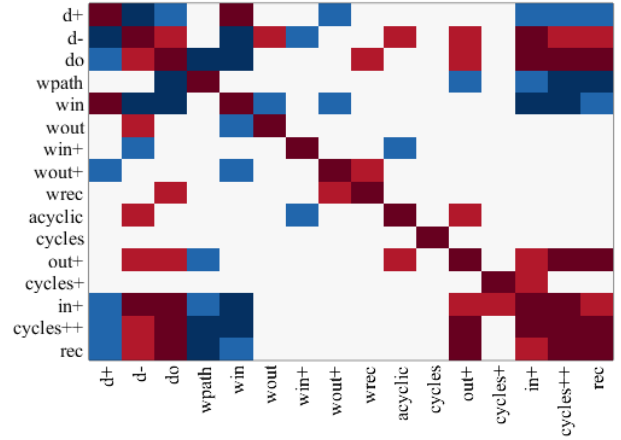


(f) Subject 12

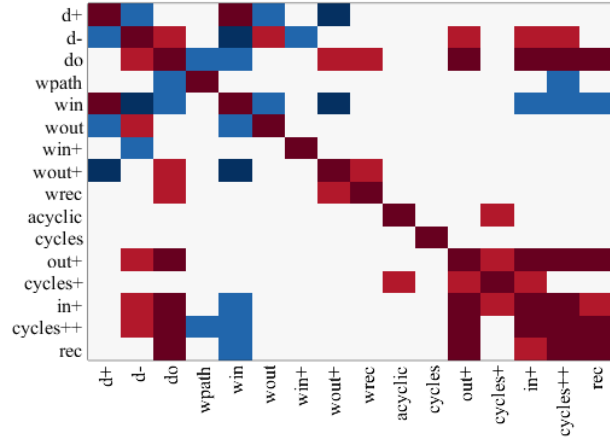
Figure 5: Graphlet correlation matrices for subjects 7 to 12



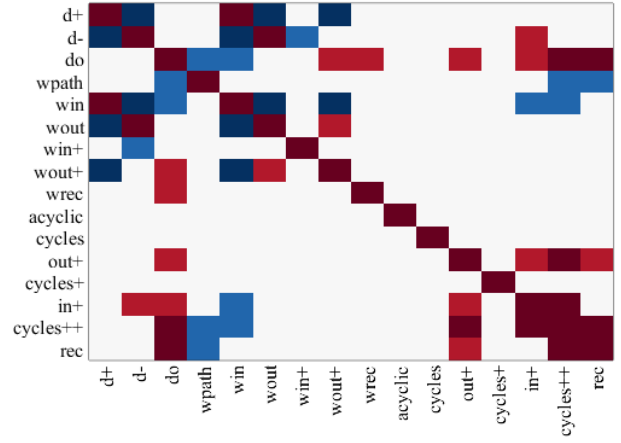
(a) Subject 13



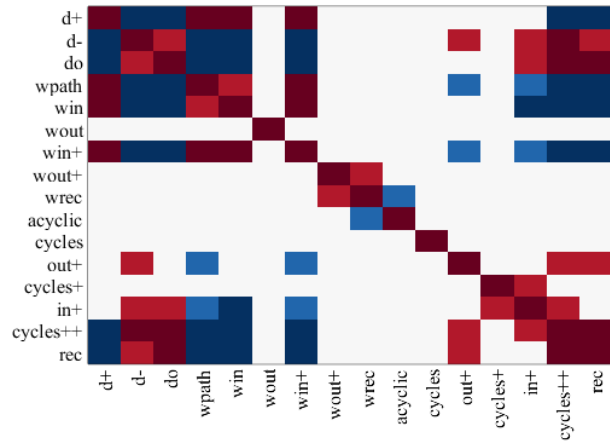
(d) Subject 16



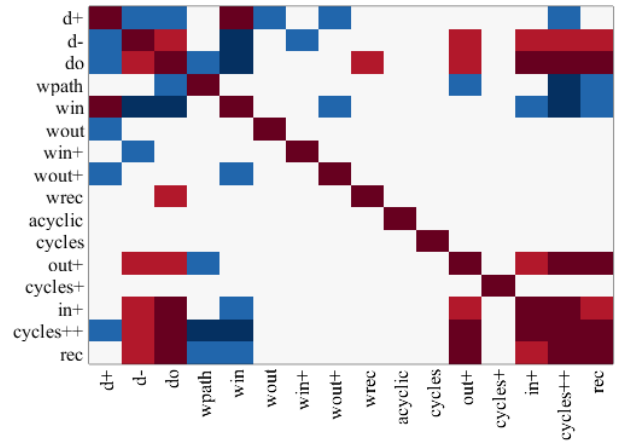
(b) Subject 14



(e) Patient 17

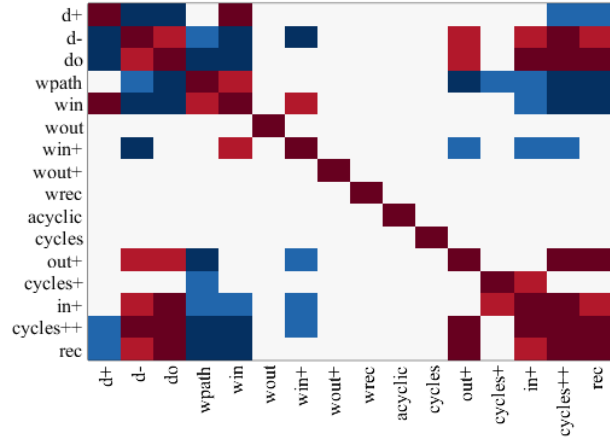


(c) Subject 15

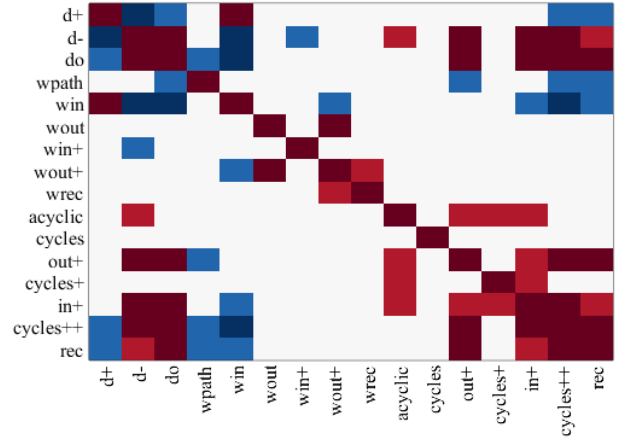


(f) Subject 18

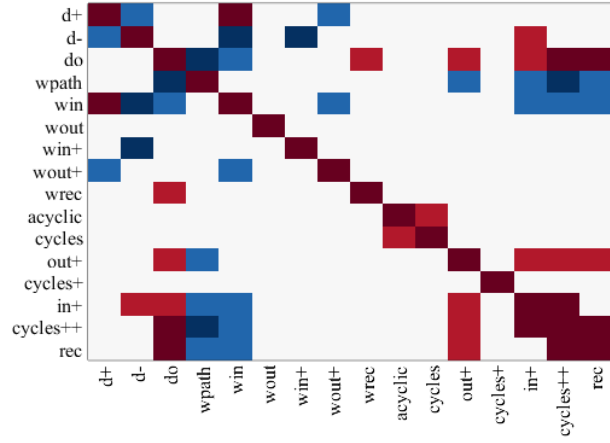
Figure 6: Graphlet correlation matrices for subjects 13 to 18



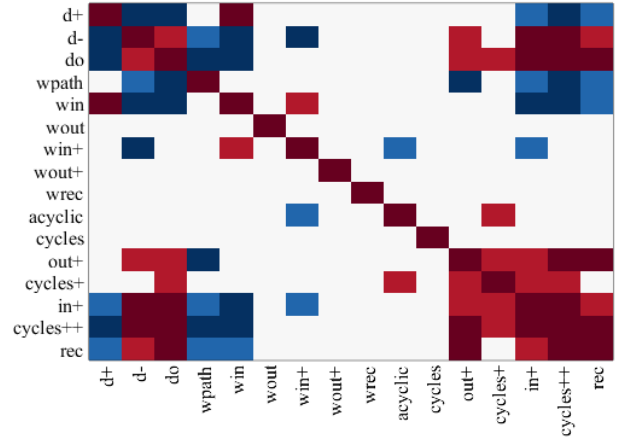
(a) Subject 19



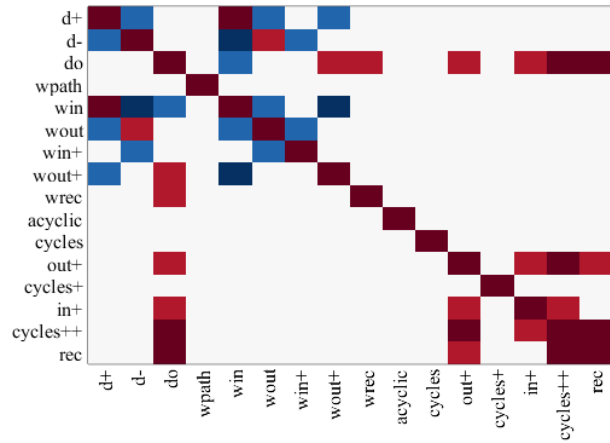
(d) Subject 22



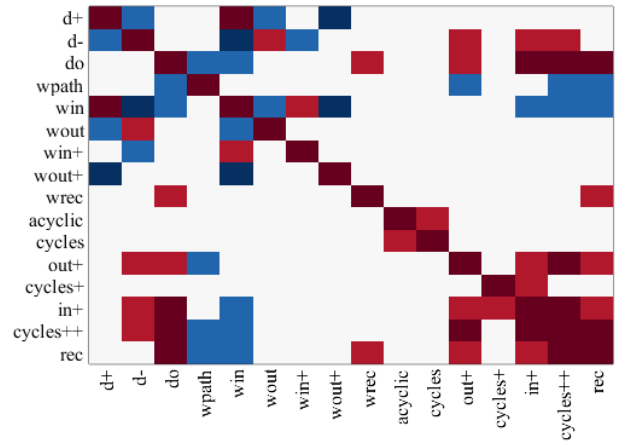
(b) Subject 20



(e) Patient 23

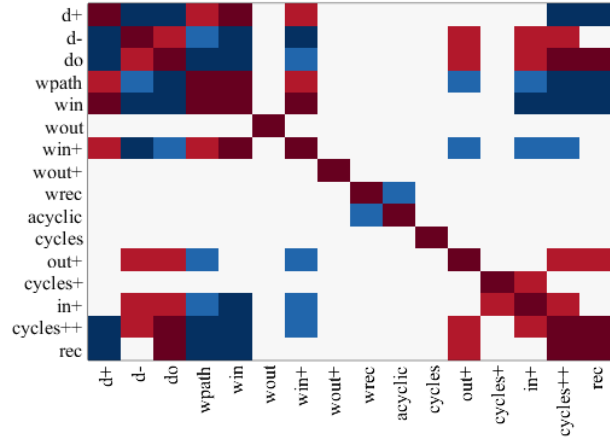


(c) Subject 22

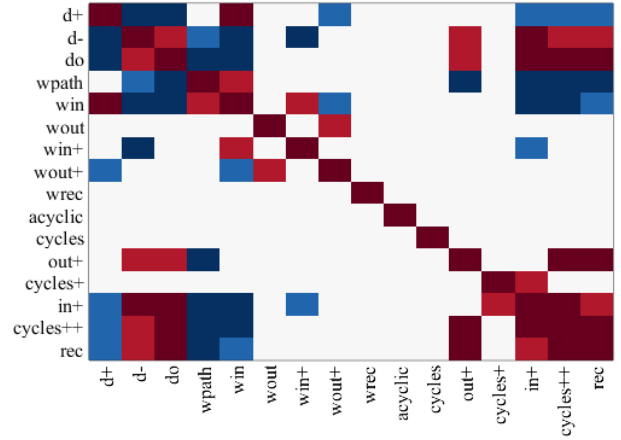


(f) Subject 24

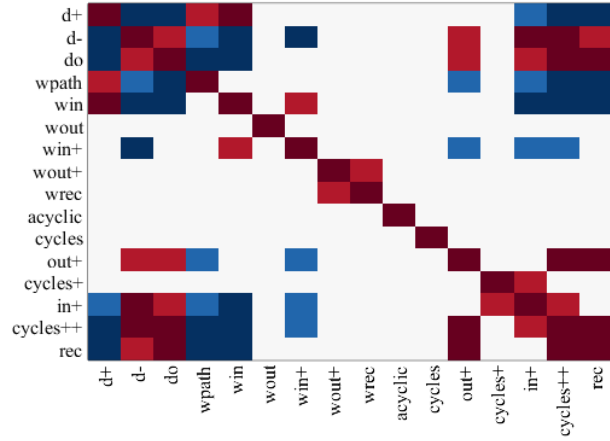
Figure 7: Graphlet correlation matrices for subjects 19 to 24



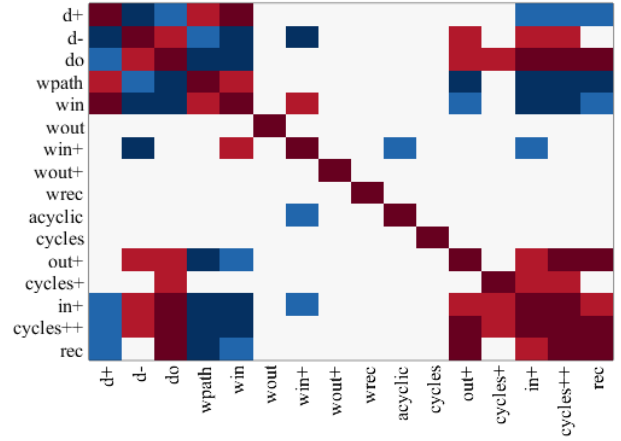
(a) Subject 25



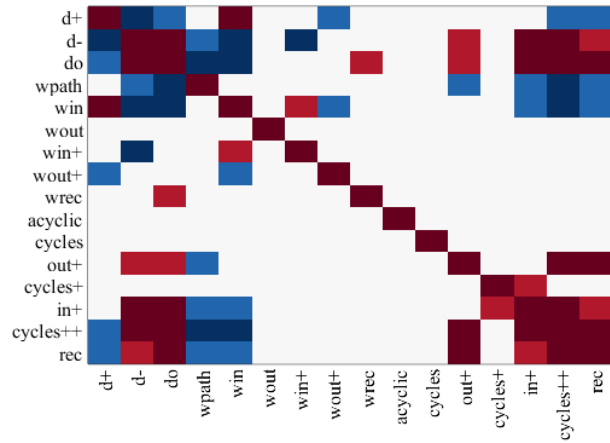
(d) Subject 28



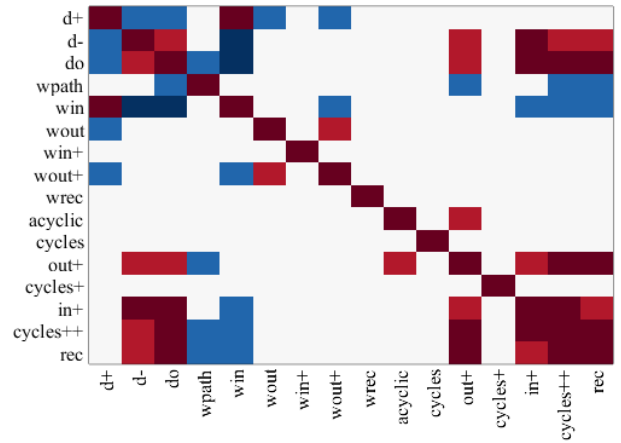
(b) Subject 26



(e) Patient 29

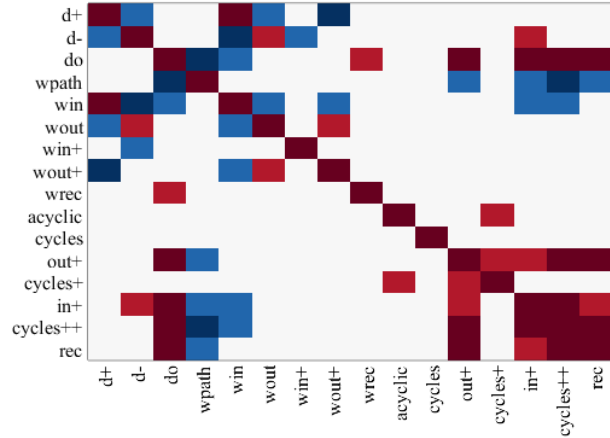


(c) Subject 27

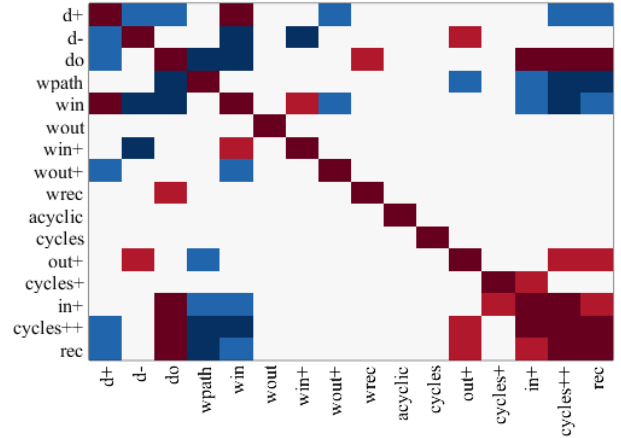


(f) Subject 30

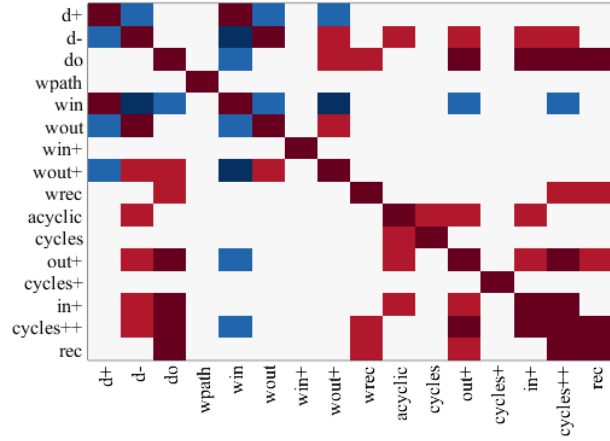
Figure 8: Graphlet correlation matrices for subjects 25 to 30



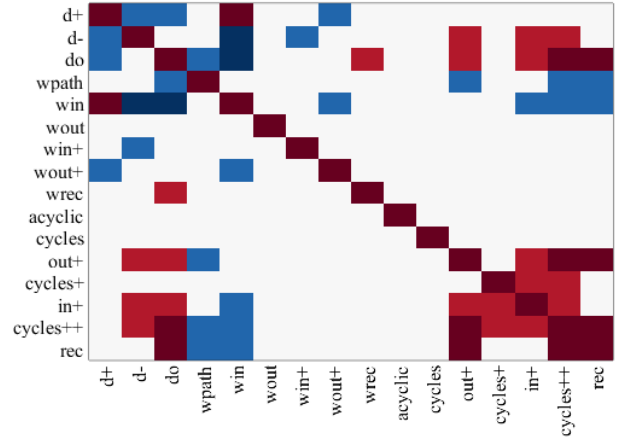
(a) Subject 31



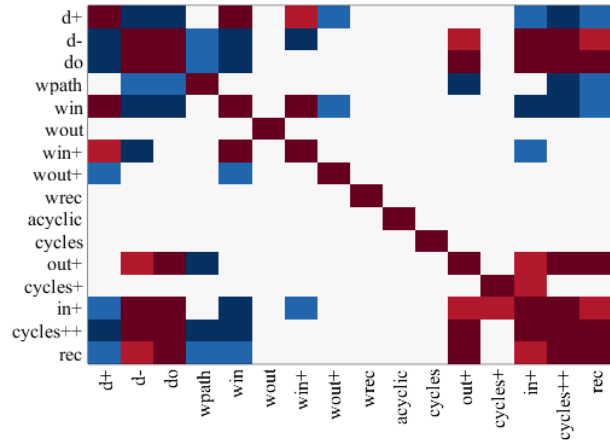
(d) Subject 34



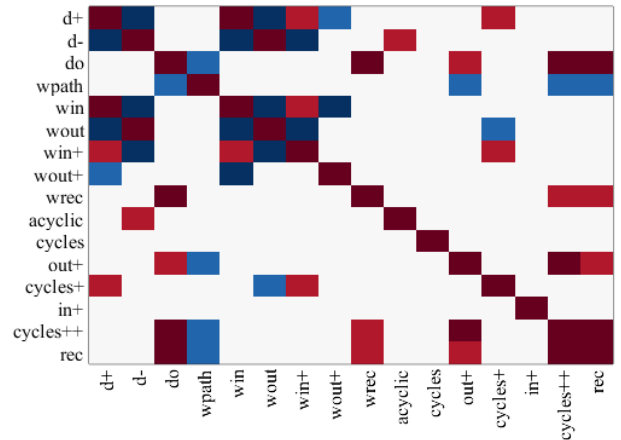
(b) Subject 32



(e) Patient 35

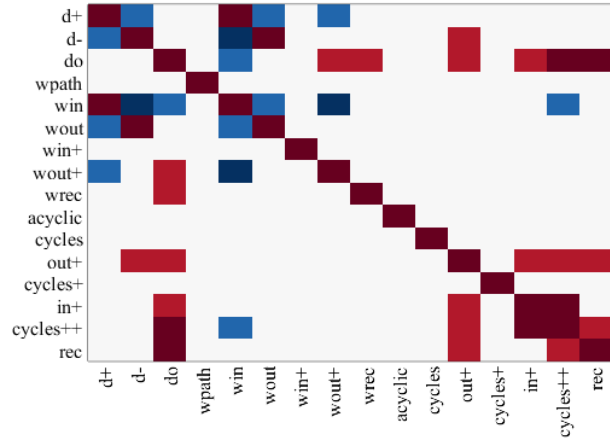


(c) Subject 33

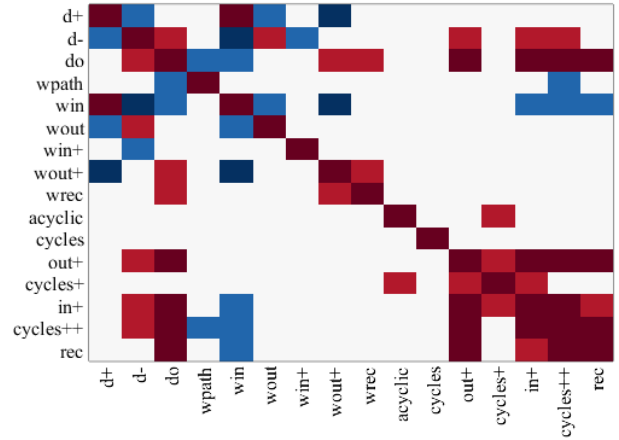


(f) Subject 36

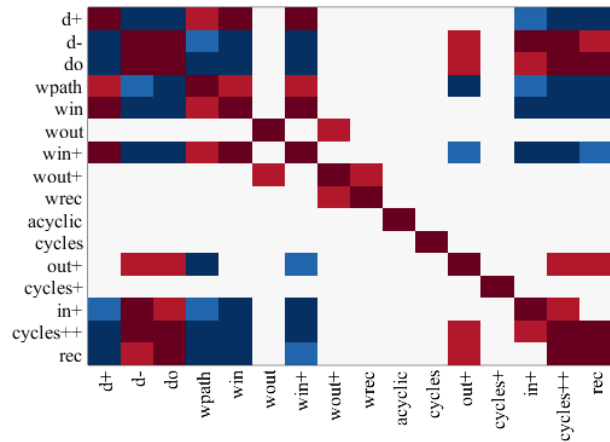
Figure 9: Graphlet correlation matrices for subjects 31 to 36



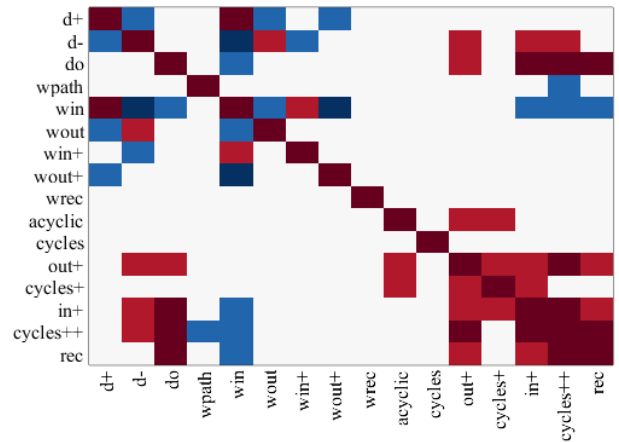
(a) Subject 37



(c) Subject 39



(b) Subject 38



(d) Patient 40

Figure 10: Graphlet correlation matrices for subjects 37 to 40

- [7] Meunier, D., Achard, S., Morcom, A., & Bullmore, E. (2009). "Age-related changes in modular organization of human brain functional networks." *Neuroimage*, 44(3), 715-723.
- [8] Zalesky, A., Fornito, A., & Bullmore, E. T. (2010). "Network-based statistic: identifying differences in brain networks.", *NeuroImage*, 53(4), 1197-207.